**Wooga Data Challenge**

The aim of this data challenge is to see your data handling and problem solving skills. We do not expect you to answer all the questions perfectly

**Task**

To optimize our spending for user acquisition one of our goals is to acquire users that have a high retention in the game. For this reason, we calculate retention rates for different cohorts. E.g. if we acquire 1000 players on day 0 and 650 players of those play the game on day 3, the day-3-retention would be 65%. If we compute this for every day we get a retention curve that shows us the retention of the cohort over time.

The given data sets contain a sample of users that installed a game in the first week of January. Instead of looking at daily retention here we are interested in weekly retention curves.

1. Exploratory Analysis
   1. Calculate and plot weekly retention curves for
      1. Different game platforms
      2. The five biggest campaigns in terms of number of installs
   2. How are the retention curves evolving over time? Which cohort has the best retention at the end of the 5-week period?
2. Data Modeling
   1. Build a simple model to fit the weekly retention rate for the five biggest campaigns (Hint: Notice the shape of the retention curves. Consider adding non-linear components). Plot fitted versus real data values for the five campaigns.
   2. How well do the model predictions fit the real data?
3. Future Predictions
   1. Use your model to predict retention rates for the weeks 6 to 10 for each of the five campaigns.
   2. Explain your results. What are your recommendations for the user acquisition team considering your findings? Do you see anything unsuspected?

**Data**

We provided you with two csv files:

* *‘user\_info.csv’* contains information on the players such as install date, game platform and country. Each row in this file corresponds to a unique player.
* *‘user\_logins.csv’* contains information on the activity of a player. Each row in this file corresponds to a different activity date of a given user. There are therefore several rows for a unique user in this file.

Data Dictionary

* *user\_id*: Unique identifier for a player.
* *install\_date*: Date when player installed the game.
* *game\_platform*: Whether the player installed on ios or android.
* campaign: Identifier of the advertising campaign through which the player installed the game. Value is missing for players that could not be attributed to a campaign.
* *activity\_date:* Date on which the player logged into / played the game.